

Beyond the Bethe Free Energy of LDPC Codes via Polymer Expansions

Nicolas Macris and Marc Vuffray
LTHC-IC-EPFL

Lausanne, Switzerland

nicolas.macris@epfl.ch, marc.vuffray@epfl.ch

Abstract—The loop series provides a formal way to write down corrections to the Bethe entropy (and/or free energy) of graphical models. We provide methods to rigorously control such expansions for low-density parity-check codes used over a highly noisy binary symmetric channel. We prove that in the asymptotic limit of large size, with high probability, the Bethe expression gives an exact formula for the entropy (per bit) of the input word conditioned on the output of the channel. Our methods also apply to more general models.

I. INTRODUCTION

Often one needs to compute the free energy and/or entropy of a graphical model. The Bethe approximation and the related Belief Propagation (BP) equations may sometimes offer a good starting point. However it is seldom a controlled approximation and even worse it is usually not clear if it yields upper or lower bounds, or even if there is any such relationship. There are not many results that precisely pinpoint the relation between the Bethe and true free energies or entropies. A general result of Vontobel [1] relates the Bethe free energy to an average of the true free energy over all graph covers. For Ising-like graphical models with attractive pair interactions, Wainwright [2] has shown that, under additional special conditions, the Bethe free energy is a bound to the true free energy. This work uses the same loop series used here. It is well known that the Bethe free energy is exact on trees, and it is natural to investigate its possible exactness on random Erdős-Rényi type graphs which are known to be locally tree-like. But we already know of systems, such as random constraint satisfaction models (e.g. K -SAT or Q -coloring) or spin glasses, where the true free energy is *not* given by the Bethe formula - even when averaged over the graph ensemble. The local tree-like nature of the graph is not sufficient when long ranged correlations are present [3].

For graphical models that describe communication with low density (parity-check and generator-matrix) codes over binary-symmetric memoryless channels the situation is favorable. Indeed we have plenty of evidence that the replica-symmetric solution¹ is exact. See [4], [5], [6] for bounds and [7], [8] for results on the binary erasure channel. In [9] it is proven that correlations between pairs of distant (with respect to Tanner graph distance) bits decay exponentially fast for LDGM codes in the regime of large noise, and LDPC codes in the regime

of small noise. This also allowed to conclude that the replica symmetric formulas are exact in these regimes.

A few years ago Chertkov and Chernyak [10] developed a loop series representation for the free energy of graphical models. The virtue of this representation is that it isolates the Bethe contribution, and represents the *remainder* by a series of terms involving only BP messages associated to *generalized loops* of the graph. It is tempting to use this representation as a tool to compare the true and Bethe free energies.

In this contribution we consider regular LDPC(l, r) codes used over a *highly noisy* BSC. Consider the conditional entropy $\frac{1}{n}H(\underline{X}|\underline{Y})$ of the input word $\underline{X} = (X_1 \cdots X_n)$ given a channel output $\underline{Y} = (Y_1 \cdots Y_n)$. We prove that in the large size limit, with high probability with respect to the code ensemble, the difference between the conditional entropy and the Bethe formula tends to zero. The error term essentially comes from the probability that the graph is not locally tree-like. Our techniques also allow to organize the dominant correction terms into a *polymer expansion*² involving generalized loops of size less than $\lambda_0 n$ ($0 < \lambda_0 < 1$ a constant). As we will show, expander arguments imply that this polymer expansion converges uniformly in n . When the terms of the polymer expansion are added to the Bethe expression, with high probability, the difference with the conditional entropy becomes $O(e^{-n\epsilon})$ for some $\epsilon > 0$.

Our results also apply to more general models. Namely the channel could have asymmetric flip probability. In fact the whole technique and results apply to spin-glass models on (l, r) Tanner graphs with l odd and $l < r$, with *small magnetic fields*, and *any temperature*. The limitation to $l < r$ is not just technical. Indeed $l > r$ would correspond to a kind of XORSAT constraint satisfaction problem, and for the usual XORSAT problem we know that the replica symmetric solutions are not generally exact at low temperatures.

The case $l = 2$ (cycle codes) has its own special features and has been discussed in [12].

II. PRELIMINARIES

We begin with a few definitions and notations. Fix two integers $l < r$. Consider two vertex sets: V a set of n *variable nodes* and C a set of $m = n \frac{l}{r}$ *check nodes*. We think of n large and l, r fixed. We consider bipartite (l, r) regular graphs

¹Replica-symmetric formulas are averaged forms of the Bethe formulas, where the average is over the channel output realizations and code ensemble.

²See [11] for a pedagogical introduction to polymer expansions.

- call them Γ - connecting V and C . The set of edges is E . More precisely, vertices of V have degree l , vertices of C have degree r , and there are no double edges. The set of all such graphs is denoted $\mathcal{B}(l, r, n)$. Note that Γ is the Tanner graph of a LDPC code with design rate $1 - l/r$. When we say that Γ is random we mean that we draw it uniformly randomly from the set $\mathcal{B}(l, r, n)$. The corresponding expectation is \mathbb{E}_Γ .

Letters i, j will always denote nodes in V and letters a, b nodes in C . We reserve the notations ∂i (resp. ∂a) for the sets of neighbors of i (resp. a) in Γ .

We will say that Γ is a (λ, κ) expander if for every subset $\mathcal{V} \subset V$ such that $|\mathcal{V}| < \lambda n$ we have $|\partial \mathcal{V}| \geq \kappa l |\mathcal{V}|$. Here $\partial \mathcal{V}$ is the number of check nodes that are connected to \mathcal{V} . Take a random Γ . We can always find $\lambda > 0$ such that with probability $1 - O(n^{-(l(1-\kappa)-1)})$, Γ is a (λ, κ) expander with $\kappa < 1 - \frac{1}{l}$. It is sufficient to take $0 < \lambda < \lambda_0$ where λ_0 is the positive solution of the equation³

$$\frac{l-1}{l} h_2(\lambda_0) - \frac{l}{r} h_2(\lambda_0 \kappa r) - \lambda_0 \kappa r h_2\left(\frac{1}{\kappa r}\right) = 0. \quad (1)$$

As will be seen later we need to take $\kappa \in [1 - \frac{2(r-1)}{lr}, 1 - \frac{1}{l}]$ (which is always possible for $r > 2$). In the rest of the paper κ is always a constant in this interval, and $0 < \lambda < \lambda_0$. For concreteness, one can take the case $(l, r) = (3, 6)$, fix $\kappa = 1/2$ and $\lambda_0 = 5 \times 10^{-4}$.

Assume that we transmit (with uniform prior) code words from an LDPC code with Tanner graph Γ over a BSC with flip probability p . We assume without loss of generality that the all zero codeword is transmitted. Then the posterior probability that $\underline{x} = (x_i)_{i=1}^n \in \{0, 1\}^n$ is the transmitted word given that $\underline{y} = (y_i)_{i=1}^n \in \{0, 1\}^n$ is received, reads

$$p_{\underline{X}|\underline{Y}}(\underline{x}|\underline{y}) = \frac{1}{Z} \prod_{a \in C} \mathbb{I}(\oplus_{i \in \partial a} x_i = 0) \prod_{i \in V} \exp((-1)^{x_i} h_i). \quad (2)$$

The graph Γ enters in this formula through the parity check constraints. In this formula $h_i = (-1)^{y_i} \frac{1}{2} \ln \frac{1-p}{p}$ and Z is the normalizing factor

$$Z = \sum_{\underline{x} \in \{0,1\}^n} \prod_{a \in C} \mathbb{I}(\oplus_{i \in \partial a} x_i = 0) \prod_{i \in V} \exp((-1)^{x_i} h_i). \quad (3)$$

We set

$$h = \frac{1}{2} \ln \frac{1-p}{p} \quad (4)$$

It is good to keep in mind that the high noise regime considered in this paper corresponds to small h (p close to $1/2$) and that $|h_i| = h$.

It is equivalent to describe the channel outputs in terms of \underline{y} or in terms of the half-log-likelihood variables $\underline{h} = (h_i)_{i=1}^n$. Note that h_i have the probability distribution $c(h_i) = (1-p)\delta(h_i - \ln \frac{1-p}{p}) + p\delta(h_i - \ln \frac{p}{1-p})$. The expectation with respect to this distribution is called $\mathbb{E}_{\underline{h}}$. We are interested in

the conditional entropy $H(\underline{X}|\underline{Y})$ of the input word given the output word. We have (see e.g., [3])

$$\mathfrak{h}_n \equiv \frac{1}{n} H(\underline{X}|\underline{Y}) = \frac{1}{n} \mathbb{E}_{\underline{h}} [\ln Z] - \frac{1-2p}{2} \ln \frac{1-p}{p}. \quad (5)$$

In (5), $n^{-1} \ln Z$ is the *free energy* of the Gibbs measure (2).

III. THE BETHE APPROXIMATION

The Bethe free energy involves a set of messages $\{\eta_{i \rightarrow a}, \hat{\eta}_{a \rightarrow i}\}$ attached to the edges of Γ . The collection of all messages is denoted $(\underline{\eta}, \underline{\hat{\eta}})$. These satisfy the BP equations

$$\begin{cases} \eta_{i \rightarrow a} &= h_i + \sum_{b \in \partial i \setminus a} \hat{\eta}_{b \rightarrow i} \\ \hat{\eta}_{a \rightarrow i} &= \tanh^{-1} \left(\prod_{j \in \partial a \setminus i} \tanh \eta_{j \rightarrow a} \right). \end{cases}$$

These equations always have a trivial solution $\tanh \eta_{i \rightarrow a} = \tanh \hat{\eta}_{a \rightarrow i} = 1$. We will consider only non-trivial solutions that are relevant for small h . For these solutions $\eta_{i \rightarrow a}$ and $\hat{\eta}_{a \rightarrow i}$ take small values and we can show that $|\eta_{i \rightarrow a}| \leq |h| + (l-1)|h|^{r-1} + O(|h|^r)$ and $|\hat{\eta}_{a \rightarrow i}| \leq |h|^{r-1} + O(|h|^r)$. We call such solutions *high-noise-solutions*.

These solutions have a Bethe free energy

$$f_{\text{Bethe}}(\underline{\eta}, \underline{\hat{\eta}}) = \frac{1}{n} \left(\sum_{a \in C} F_a + \sum_{i \in V} F_i - \sum_{(i,a) \in E} F_{ia} \right), \quad (6)$$

where

$$\begin{cases} F_a = \ln \frac{1}{2} (1 + \prod_{i \in \partial a} \tanh \eta_{i \rightarrow a}) + \sum_{i \in \partial a} \ln 2 \cosh \eta_{i \rightarrow a}, \\ F_i = \ln 2 \cosh \left(h_i + \sum_{a \in \partial i} \hat{\eta}_{a \rightarrow i} \right), \\ F_{ia} = \ln 2 \cosh (\eta_{i \rightarrow a} + \hat{\eta}_{a \rightarrow i}). \end{cases}$$

Theorem 1: Suppose l is odd and $3 \leq l \leq r$. There exists $h_0 > 0$ (small) independent of n , such that for $|h| \leq h_0$ and any high-noise-solution $(\underline{\eta}, \underline{\hat{\eta}})$ of the BP equations,

$$\mathbb{E}_\Gamma \left[\left| \frac{1}{n} \ln Z - f_{\text{Bethe}}(\underline{\eta}, \underline{\hat{\eta}}) \right| \right] = O\left(\frac{1}{n^{l(1-\kappa)-1}}\right). \quad (7)$$

The $O(\cdot)$ is uniform in the channel output realizations \underline{h} .

Remark 1: By Markov's bound we obtain that the difference between the true and Bethe free energies tends to zero with high probability, in the $n \rightarrow +\infty$ limit.

Remark 2: We can average equation (7) over the channel output and use (5) to relate the true and Bethe entropies.

IV. LOOP CORRECTIONS TO THE BETHE APPROXIMATION

We define a *generalized loop* g as any subgraph contained in Γ with no dangling edges (figure 1). Note that a generalized loop is not necessarily connected. We call $d_i(g)$ (resp. $d_a(g)$) the *induced degree* of node i (resp. a) in g . For a generalized loop we have $d_i(g) \in \{2, \dots, l\}$ and $d_a(g) \in \{2, \dots, r\}$.

For a finite size system, the *loop series* [10] is an identity valid for any solution of the BP equations. We have

$$\frac{1}{n} \ln Z - f_{\text{Bethe}}(\underline{\eta}, \underline{\hat{\eta}}) = \frac{1}{n} \ln \left\{ \sum_{g \subset \Gamma} K(g) \right\}. \quad (8)$$

The sum on the right hand side carries over all generalized loops included in Γ . The $K(g)$ can be expressed entirely in

³See e.g [7] where the standard LDPC(l, r, n) ensemble is considered. It is easily argued that the same result applies to $\mathcal{B}(l, r, n)$.

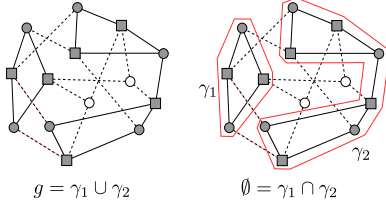


Fig. 1. Example of $\Gamma \in \mathcal{B}(3, 4, 8)$. The generalized loop g has two disjoint connected parts γ_1 and γ_2 .

terms of BP messages $\eta_{i \rightarrow a}$ and $\hat{\eta}_{a \rightarrow i}$. The explicit formula is given in the appendix. Remarkably $K(g)$ factorizes in a product of contributions associated to the connected parts of g . Each generalized loop can be decomposed in a unique way as a union $g = \cup_k \gamma_k$ where γ_k are *connected and disjoint generalized loops*. The γ_k 's are called *polymers*. We have $K(g) = \prod_k K(\gamma_k)$ and

$$\sum_{g \subset \Gamma} K(g) = \sum_{M \geq 0} \frac{1}{M!} \sum_{\gamma_1, \dots, \gamma_M \subset \Gamma} \prod_{k=1}^M K(\gamma_k) \times \prod_{k < k'} \mathbb{I}(\gamma_k \cap \gamma_{k'} = \emptyset). \quad (9)$$

In the sum each γ_k runs over all polymers contained in Γ . The factor $\frac{1}{M!}$ accounts for the fact that a polymer configuration has to be counted only once. Finally the indicator function ensures that the polymers do not intersect. Because of this constraint all sums in (9) are finite.

From a physical point of view (9) is the partition function of polymers that can acquire any shape allowed by Γ , have *activity*⁴ $K(\gamma)$, and interact via a two body hard-core repulsion. This analogy allows us to use methods from statistical mechanics to analyze the corrections to the Bethe free energy.

We say that a *polymer is small* if $|\gamma| < \lambda n$ for some fixed λ that we take in the interval $[0, \lambda_0]$. The contribution of small polymers to (9) is

$$Z_p(\underline{\eta}, \hat{\underline{\eta}}) = \sum_{M \geq 0} \frac{1}{M!} \sum_{\gamma_1, \dots, \gamma_M \text{ s.t. } |\gamma_k| < \lambda n} \prod_{k=1}^M K(\gamma_k) \prod_{k < k'} \mathbb{I}(\gamma_k \cap \gamma_{k'} = \emptyset). \quad (10)$$

Theorem 2: Suppose l is odd and $3 \leq l \leq r$. take Γ at random. There exist a small h_0 independent of n such that for $|h| < h_0$, and any high-noise-solution $(\underline{\eta}, \hat{\underline{\eta}})$ of the BP equations, with probability $1 - \frac{1}{\epsilon} O(n^{-(l(1-\kappa)-1)})$,

$$\frac{1}{n} \ln Z = f_{\text{Bethe}}(\underline{\eta}, \hat{\underline{\eta}}) + \frac{1}{n} \ln Z_p + O(e^{-\epsilon n}) \quad (11)$$

for $\epsilon > 0$. Here $O(\cdot)$ is uniform \underline{h} .

The second term on the right hand side of (11) is the partition function of small polymers. One can compute in a systematic way the leading corrections to the Bethe free energy

⁴This is the name used by chemists to denote the probability weight assuming that the polymer would be isolated. Note that here $K(\gamma)$ can be negative and this analogy is at best formal.

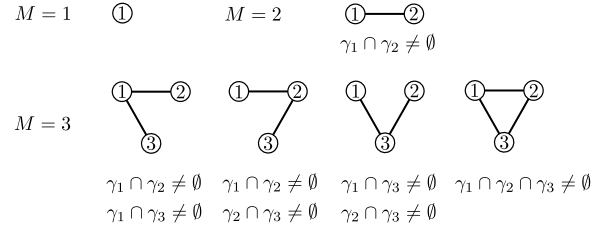


Fig. 2. All the Mayer graphs for $M = 1, 2, 3$.

by expanding the logarithm in powers of the activities $K(\gamma)$. This yields the so-called *polymer (or Mayer) expansion*,

$$\frac{1}{n} \ln Z_p(\underline{\eta}) = \frac{1}{n} \sum_{M \geq 1} \frac{1}{M!} \sum_{\gamma_1, \dots, \gamma_M \text{ s.t. } |\gamma_k| < \lambda n} \prod_{k=1}^M K(\gamma_k) \times \sum_{G \subset \mathcal{G}_M} \prod_{(k, k') \in G} (-\mathbb{I}(\gamma_k \cap \gamma_{k'} \neq \emptyset)). \quad (12)$$

The third sum is over the set \mathcal{G}_M of all connected *Mayer graphs* G with M vertices labeled by $\gamma_1, \dots, \gamma_M$ (see figure 2). Note that in the expansion of the logarithm, the indicator function forces the polymers to overlap. Therefore the summations contains an infinite number of terms and its convergence has to be controlled.

Lemma 1: Suppose $r > 2$. Fix $\zeta_0 > 1$ and replace $K(\gamma)$ by $\zeta K(\gamma)$ ($\zeta \in \mathbb{C}$) in the polymer expansion (12) which then becomes a power series in the parameter $|\zeta| \leq \zeta_0$. Assume that Γ is a (λ, κ) expander with $\kappa \in [1 - \frac{2(r-1)}{lr}, 1 - \frac{1}{l}]$. One can find $h_0 > 0$ such that for $|h| < h_0$ this power series is absolutely convergent uniformly in n and \underline{h} .

Remark 3: This lemma holds for any (l, r) with $r > 2$.

Remark 4: Our real interest is of course for $\zeta = 1$, and the introduction of the parameter ζ above is just a convenient way to describe the nature of the polymer expansion. The lemma implies that one can compute the limit $n \rightarrow +\infty$ of the polymer expansion *term by term* (for small polymers), and that this limit is analytic for $|\zeta| < \zeta_0$. This lemma forms a crucial part for the proofs of theorems 1 and 2.

Remark 5: The last term in the right hand side of (11) contains the contributions of *large* polymers of size greater than λn (in a sea of small polymers). It turns out that this contribution cannot be expanded into an absolutely convergent series, and has to be treated non-perturbatively by counting methods.

Lemma 1 has the following consequence:

Corollary 1: Suppose $r > 2$. One can find $h_0 > 0$ independent of n such that for $|h| < h_0$,

$$\frac{1}{n} \mathbb{E}_\Gamma [\ln Z_p(\underline{\eta}, \hat{\underline{\eta}})] = O\left(\frac{1}{n^{l(1-\kappa)-1}}\right) \quad (13)$$

V. CONVERGENCE OF THE POLYMER EXPANSION (12)

We give the main ideas of the proof of lemma 1.

Proof of Lemma 1: A standard criterion for uniform

convergence and analyticity of the polymer expansion is [11]

$$Q \equiv \sum_{t=0}^{\infty} \frac{1}{t!} \sup_{z \in V \cup C} \sum_{\gamma \ni z, |\gamma| < \lambda n} |\gamma|^t \zeta_0 |K(\gamma)| < 1. \quad (14)$$

If we prove that for polymers such that $|\gamma| < \lambda n$ we have

$$|K(\gamma)| \leq h^{\frac{r}{2}|\gamma|}, \quad (15)$$

then the result follows for h small enough.

The main difficulty in proving (15) is that the (optimal) estimate (34), (35) in the Appendix shows that $K(\gamma)$ is not necessarily very small for graphs containing too many check nodes of maximal induced degree and too many variable nodes of even induced degree. More precisely for these bad graphs the activity is not exponentially small in the size of the graph. Then it is not possible to compensate for the "entropy" of the graph.

We will use an expander argument to show that these bad cases *do not occur* when $|\gamma| < \lambda n$. We derive (15) with

$$c = r - \frac{2+r}{3-l(1-\kappa)}. \quad (16)$$

In the process of this derivation one has to require $3-l(1-\kappa) > 0$ and $c > 0$. This imposes the condition on the expansion constant $\kappa > 1 - \frac{2(r-1)}{lr}$. Note that an expansion constant cannot be greater than $1 - 1/l$, so it is fortunate that we have $1 - \frac{1}{l} > 1 - \frac{2(r-1)}{lr}$ (for any $r > 2$).

Now we sketch the proof of (15) and (16). Recall that $d_i(\gamma)$ (resp. $d_a(\gamma)$) is the induced degree of node i (resp. a) in γ . The type of γ is given by two vectors $\underline{n} = (n_s(\gamma))_{s=2}^l$ and $\underline{m} = (m_t(\gamma))_{t=2}^r$ defined as $n_s(\gamma) := |\{i \in \gamma \cap V | d_i(\gamma) = s\}|$ and $m_t(\gamma) := |\{a \in \gamma \cap C | d_a(\gamma) = t\}|$. In words, $n_s(\gamma)$ and $m_t(\gamma)$ count the number of variable and check nodes with induced degrees s and t in γ . Note that we have the constraints

$$\begin{cases} |\gamma| = \sum_{s=2}^l n_s(\gamma) + \sum_{t=2}^r m_t(\gamma) \\ \sum_{s=2}^l s n_s(\gamma) = \sum_{t=2}^r t m_t(\gamma) \end{cases} \quad (17)$$

We apply the expander property to the set $\mathcal{V} = \{i \in \gamma \cap V\}$. This reads

$$|\partial \mathcal{V}| \geq \kappa l \left| \sum_{s=2}^l n_s(\gamma) \right|. \quad (18)$$

On the other hand $|\partial \mathcal{V}| \leq \sum_{t=2}^{r-1} m_t(\gamma) + \sum_{s=2}^l (l-s) n_s(\gamma)$. With (18) this yields the constraint

$$\sum_{t=2}^{r-1} m_t(\gamma) + \sum_{s=2}^l (l-s) n_s(\gamma) \geq \kappa l \left| \sum_{s=2}^l n_s(\gamma) \right|. \quad (19)$$

Using all constraints (17) and (19) we can prove

$$\sum_{t=2}^{r-1} (r-t) m_t(\gamma) \geq \left(r - \frac{2+r}{3-l(1-\kappa)} \right) |\gamma|. \quad (20)$$

Finally, keeping only the product over $t = 2, \dots, r-1$ in estimates (34) and (35) in the Appendix, we obtain (15). ■

Proof of Corollary 1: Conditional on Γ being an expander we have from the previous proof $0 < Q < 1$. Then,

polymer expansion techniques [11] allow to estimate the sum over M in (12) term by term, which yields

$$\left| \frac{1}{n} \ln Z_p(\underline{n}, \underline{m}) \right| \leq (1-Q)^{-1} n^{-1} \sum_{z \in V \cup C} \sum_{\gamma \ni z, |\gamma| < \lambda n} |K(\gamma)| e^{|\gamma|}. \quad (21)$$

If we take the expectation over graphs we cancel the sum over $z \in V \cup C$ and the n^{-1} . This allows to consider a sum of polymers rooted at one vertex. We compute this expectation by conditioning on the *first* event that Γ is tree-like in a neighborhood of size $O(\ln n)$ around this vertex, and on the *second* complementary event. The second event has small probability $O(n^{-(1-\beta)})$ for any $0 < \beta < 1$. Besides from (21) and (15) it is easy to show that $n^{-1} |\ln Z_p|$ is bounded. For the first event we have that the smallest polymer is a cycle with $|\gamma| = O(\ln n)$. This with (21) and (15) implies that $n^{-1} |\ln Z_p(\underline{n}, \underline{m})| \leq n^{-\beta |\ln |h||}$. Combining all these remarks with the fact that Γ is an expander with probability $1 - O(n^{-(l(1-\kappa)-1)})$ we obtain (13). ■

VI. PROBABILITY ESTIMATES ON GRAPHS

In this section we deal with the contribution $R(\underline{n}, \underline{m})$ corresponding to terms containing at least one large polymer in (9). We have

$$\sum_{g \subset \Gamma} K(g) = Z_p(\underline{n}, \underline{m}) + R(\underline{n}, \underline{m}), \quad (22)$$

where

$$R(\underline{n}, \underline{m}) = \sum_{g \subset \Gamma \text{ s.t. } \exists \gamma \subset g \text{ with } |\gamma| \geq \lambda n} K(g), \quad (23)$$

The next lemma shows that the contribution from large polymers is exponentially small, with high probability with respect to the graph ensemble.

Lemma 2: Fix $\delta > 0$. Assume $l \geq 3$ odd and $l < r$. There exists a constant $C > 0$ depending only on l and r such that for h small enough

$$\mathbb{P} [|R(\underline{n}, \underline{m})| \geq \delta] \leq \frac{1}{\delta} e^{-Cn} \quad (24)$$

Sketch of Proof: Let $\Omega_{\Gamma}(\underline{n}, \underline{m})$ be the set of all $g \subset \Gamma$ with prescribed type $(\underline{n}(g), \underline{m}(g))$. By (35) and the Markov bound

$$\begin{aligned} \mathbb{P} \left[\sum_{g \subset \Gamma \text{ with } |g| \geq \lambda n} |K(g)| \geq \delta \right] \\ \leq \frac{1}{\delta} \sum_{\vec{n}, \vec{m} \in \Delta} \bar{K}(\vec{n}, \vec{m}) \mathbb{E}_{\Gamma} [|\Omega_{\Gamma}(\vec{n}, \vec{m})|], \end{aligned} \quad (25)$$

Notice that the probability in (25) is an upper bound on the probability in (24). In (25) we have

$$\begin{aligned} \Delta \equiv \left\{ (\underline{n}, \underline{m}) \mid \lambda n \leq \sum_{s=2}^l n_s + \sum_{t=2}^r m_t, \sum_{s=2}^l s n_s = \sum_{t=2}^r t m_t, \right. \\ \left. \sum_{s=2}^l n_s < n, \sum_{t=2}^r m_t < nl/r \right\}. \end{aligned} \quad (26)$$

The expectation of the number of $g \subset \Gamma$ with prescribed type can be estimated by combinatorial bounds provided by McKay [13]. It turns out that these subgraphs proliferate exponentially in n only for a subdomain of Δ where $\overline{K}(\underline{n}, \underline{m})$ is exponentially smaller in n . In the subdomain where $\overline{K}(\underline{n}, \underline{m})$ is not small (but it is always bounded) the number of subgraphs is subexponential when l is odd and $l < r$. As a consequence for l odd and $l < r$, we are able to prove that the sum on the right hand side of (25) is smaller than e^{-Cn} . Unfortunately our estimates break down for l even. ■

VII. SKETCH OF PROOF OF THEOREMS 1 AND 2

We write

$$\frac{1}{n} \ln \left\{ \sum_{g \subset \Gamma} K(g) \right\} = \frac{1}{n} \ln Z_p(\underline{\eta}, \underline{\hat{\eta}}) + \frac{1}{n} \ln \left(1 + \frac{R(\underline{\eta}, \underline{\hat{\eta}})}{Z_p(\underline{\eta}, \underline{\hat{\eta}})} \right). \quad (27)$$

We first look at the second contribution coming from large polymers. From corollary 1 and the Markov bound, we have for any $\epsilon > 0$,

$$\mathbb{P}[e^{-n\epsilon} \leq \frac{1}{Z_p(\underline{\eta}, \underline{\hat{\eta}})} \leq e^{n\epsilon}] = 1 - \frac{1}{\epsilon} O(n^{-(l(1-\kappa)-1)}) \quad (28)$$

Using inequalities (24) and (28), and choosing $\delta = e^{-2n\epsilon}$ it is not difficult to show that (at this point one takes $2\epsilon < C$)

$$\mathbb{P} \left[\left| \frac{R(\underline{\eta}, \underline{\hat{\eta}})}{Z_p(\underline{\eta}, \underline{\hat{\eta}})} \right| \geq e^{-n\epsilon} \right] \leq \frac{1}{\epsilon} O(n^{-(l(1-\kappa)-1)}) + e^{-n(C-2\epsilon)}. \quad (29)$$

This allows to conclude that with probability $1 - \frac{1}{\epsilon} O(n^{-(l(1-\kappa)-1)})$

$$\frac{1}{n} \ln \left(1 + \frac{R(\underline{\eta}, \underline{\hat{\eta}})}{Z_p(\underline{\eta}, \underline{\hat{\eta}})} \right) = O(e^{-n\epsilon}). \quad (30)$$

This already proves theorem 2.

It is now easy to show theorem 1. There is a probability $O(n^{-(l(1-\kappa)-1)})$ that this last term is not small. However we can always show it is bounded by a constant independent of n . Indeed it is equal to the difference $n^{-1} \ln Z - f_{\text{Bethe}}(\underline{\eta}, \underline{\hat{\eta}}) - n^{-1} \ln Z_p(\underline{\eta}, \underline{\hat{\eta}})$ where each term separately can be shown to be bounded by a constant independent of n . Furthermore, corollary 1 tells us that the expectation of the absolute value of the first term on the r.h.s is $O(n^{-(l(1-\kappa)-1)})$. Combining these remarks allows to conclude the proof of theorem 1.

VIII. APPENDIX

We have

$$K(g) = \prod_{i \in g \cap V} K_i \prod_{a \in g \cap C} K_a \quad (31)$$

Quantities K_a, K_i are local and can be computed only with BP messages. Let $m_i = \tanh(h_i + \sum_{a \in \partial i} \hat{\eta}_{a \rightarrow i})$.

$$K_i = \frac{(1 - m_i)^{d_i(g)-1} + (-1)^{d_i(g)-1} (1 + m_i)^{d_i(g)-1}}{2(1 - m_i^2)^{d_i(g)-1}} \quad (32)$$

$$K_a = \prod_{i \in \partial a \cap g} \sqrt{\frac{1 - \tanh^2 \eta_{i \rightarrow a}}{1 - \prod_{j \in \partial a \setminus i} \tanh^2 \eta_{j \rightarrow a}}} \prod_{i \in \partial a \cap g^c} \tanh \eta_{i \rightarrow a} \\ \times \frac{1 + (-1)^{d_a(g)} \prod_{i \in \partial a} \tanh^{d_a(g)-1} \eta_{i \rightarrow a}}{1 + \prod_{i \in \partial a} \tanh \eta_{i \rightarrow a}} \prod_{i \in \partial a \cap g} \sqrt{1 - m_i^2} \quad (33)$$

Using these formulas and the BP equations we derive the following estimate for $|h_i| < h_0$ small enough

$$|K(g)| \leq \overline{K}(\underline{n}(g), \underline{m}(g)) \quad (34)$$

where

$$\overline{K}(\underline{n}(g), \underline{m}(g)) = (1 - \alpha_r r h^2)^{m_r(g)} \prod_{t=2}^{r-1} (\alpha_t h^{r-t})^{m_t(g)} \\ \times \prod_{\substack{s=2, \\ \text{even}}}^{l-1} \left(1 + \frac{\beta_s}{2} s(s-1) h^2 \right)^{n_s(g)} \prod_{\substack{s=3, \\ \text{odd}}}^l (\beta_s (s-1) h)^{n_s(g)}. \quad (35)$$

Here $0 < \alpha_r < 1$, $\alpha_t > 1$, $\beta_t > 1$ are fixed numerical constants (that we can take close to 1). Estimate (35) is essentially optimal for small h as can be checked by Taylor expanding $K(g)$ in powers of h_i .

Acknowledgment. The work of M.V was supported by the Swiss National Science Foundation grant no 200021-121903.

REFERENCES

- [1] P. O. Vontobel, *Counting in graph covers: a combinatorial characterization of the Bethe entropy function*, available at arXiv:1012.0065v1.
- [2] E. B. Sudderth, M. J. Wainwright and A. S. Willsky, *Loop series and Bethe variational bounds in attractive graphical models*, Proceedings of Neural Information Processing (2007).
- [3] M. Mézard, A. Montanari, *Information, Physics, and Computation*, Oxford University Press (2009).
- [4] A. Montanari, *Tight bounds for LDPC and LDGM codes under MAP decoding*, IEEE Trans. Inf. Theory, vol 51 pp. 3221 - 3246 (2005).
- [5] N. Macris, *Griffith-Kelly-Sherman Correlation Inequalities: A Useful Tool in the Theory of Error Correcting Codes*, IEEE Trans. Inf. Theory, vol 53 pp. 664 - 683 (2007).
- [6] S. Kudekar, N. Macris, *Sharp Bounds for Optimal Decoding of Low-Density Parity-Check Codes*, IEEE Trans. Inf. Theory, vol 55 pp. 4635-4650 (2009).
- [7] T. Richardson, R. Urbanke, *Modern Coding Theory*, Cambridge University Press, (2008).
- [8] S. Korada, N. Macris, S. Kudekar, *exact solution for the conditional entropy of Poissonian LDPC codes over the Binary erasure Channel*, in Proc. IEEE. Int. Symp. Inf. Theory pp. 1016-1020 (2007).
- [9] S. Kudekar, N. Macris, *Decay of Correlations for Sparse Graph Error Correcting Codes*, SIAM J. Discrete Math. 25(2) pp. 956-988 (2011).
- [10] M. Chertkov, V. Chernyak, *Loop series for discrete statistical models on graphs*, J. Stat. Mech., 1-28, P06009, (2006).
- [11] D. Brydges, *A short course on cluster expansions*, in K. Osterwalder and R. Stora ed. Les Houches, session XLIII, Part I, (1984).
- [12] N. Macris, M. Vuffray, *Polymer Expansions for Cycle LDPC Codes*, International Zurich Seminar on Communications, Zürich, Switzerland, (2012).
- [13] B.D. McKay, *Subgraphs of random graphs with specified degrees*, Proceedings of the International Congress of Mathematicians, Hyderabad, India, (2010).